

Секция 3: Средства создания и поддержки проблемно-ориентированных систем,
основанных на знаниях, и экспертных систем

C 10.16.1.0/30 is directly connected, Serial0/0
O 10.16.1.4/30 [110/65] via 10.16.1.10, 00:03:57, FastEthernet0/0
O IA 10.32.0.0/28 [110/129] via 10.16.1.1, 00:02:16, Serial0/0

Протокол маршрутизации OSPF относится к протоколу маршрутизации по состоянию канала, и он обладает точной информацией о топологии сети передачи данных, исходя из которой, они строят таблицу маршрутизации [2].

При применении полностью тупиковой зоны в таблице маршрутизации присутствуют только маршруты до маршрутизаторов, находящихся внутри этой зоны и маршрут по умолчанию. Даже на таком небольшом стенде произошло значительное уменьшение таблицы маршрутизации, при увеличении количества маршрутизаторов, не относящихся к полностью тупиковой зоне, разница между таблицами топологии и занимаемой ими памятью маршрутизаторов, находящихся в полностью тупиковой зоне, и находящихся в общем домене, без разделения на зоны будет только возрастать.

При создании полностью тупиковых зон для всей сети передачи данных Красноярской железной дороги каждый из маршрутизаторов будет вынужден хранить информацию не обо всех 9153 сетях, а в среднем о 610 сетях. Таким образом, в среднем будет задействовано количество памяти, приведенное в формуле 1.

$$610 \cdot 136 = 82\,960 \text{ байт} = 81.016 \text{ Кба} \quad (1)$$

Информация о количестве памяти, занимаемой записью об одном маршруте, взята из таблицы, где оно было подсчитано с проведением моделирования на стендах с различным количеством сетей. Для хранения записей обо всех маршрутах необходимо 1215.63 Кбайт памяти.

Из расчётов становится понятно, что применение полностью тупиковых зон позволяет значительно уменьшить объем занимаемой памяти и снизить нагрузку на процессор, который вынужден обрабатывать эту информацию.

Литература.

1. Bouillard, A., Jard, C., Junier, A. The impact of initial delays in ospf routing. – Communication in Computer and Information Science. – 456 – 2014: 37-57.
2. Guo, Y., Wang, Z., Yin, X., Shi, X., Wu, J. Traffic engineering in SDN/OSPF hybrid network. – Proceedings International Conference on Network Protocols. – 22nd IEEE International Conference on Network Protocols – 2014: 563-568.
3. Makarov, S.V., Ostanin, V.V., and Vaitkov, I.V. The comparison of routers by Firms Cisco, Juniper and Huawei. – Modern Applied Science. – 8(6) – 2014.
4. Rosenberg, E. and Uttaro, J. A fast re-route method. – IEEE Communications Letters – 17(8) – 2013: 1656-1659.

ПЛЮСЫ И МИНУСЫ ПРИМЕНЕНИЯ BIG DATA В ЭЛЕКТРОННОЙ КОММЕРЦИИ

И.А. Валентов, студ.

Научный руководитель: Захарова А.А. к.т.н.

Юргинский технологический институт (филиал) Национального исследовательского

Томского политехнического университета

652055, Кемеровская обл., г. Юрга, ул. Ленинградская, 26, тел. (38451) 7-77-64

E-mail: valentov-2008@rambler.ru

Big Data, «Большие данные» вот уже несколько лет как стали притчей во языцех в IT-и маркетинговой прессе. И понятно: цифровые технологии пронизали жизнь современного человека, «все пишется». Объем данных о самых разных сторонах жизни растет, и одновременно растут возможности хранения информации.

Большинство экспертов сходятся во мнении, что ускорение роста объема данных является объективной реальностью. Социальные сети, мобильные устройства, данные с измерительных устройств, бизнес-информация – вот лишь несколько видов источников, способных генерировать гигантские объемы информации. По данным исследования IDC Digital Universe, опубликованного в 2012 году, ближайшие 8 лет количество данных в мире достигнет 40 Зб (zettabytes) что эквивалентно 5200 Гб на каждого жителя планеты.

Парадигма Big Data определяет три основных типа задач.

- Хранение и управление объемом данных в сотни терабайт или петабайт, которые обычные реляционные базы данных не позволяют эффективно использовать.
- Организация неструктурированной информации, состоящей из текстов, изображений, видео и других типов данных.
- Анализ Big Data, который ставит вопрос о способах работы с неструктурированной информацией, генерацию аналитических отчетов, а также внедрение прогностических моделей.[1]

Рынок проектов Big Data пересекается с рынком бизнес-аналитики (БА), объем которого в мире, по оценкам экспертов, в 2012 году составил около 100 млрд. долларов. Он включает в себя компоненты сетевых технологий, серверов, программного обеспечения и технических услуг.

Также использование технологий Big Data актуально для решений класса гарантирования доходов (РА), предназначенных для автоматизации деятельности компаний. Современные системы гарантирования доходов включают в себя инструменты обнаружения несоответствий и углубленного анализа данных, позволяющие своевременно обнаружить возможные потери, либо искажение информации, способные привести к снижению финансовых результатов. На этом фоне российские компании, подтверждающие наличие спроса технологий Big Data на отечественном рынке, отмечают, что факторами, которые стимулируют развитие Big Data в России, являются рост данных, ускорение принятия управленческих решений и повышение их качества.

Сегодня анализируется только 0,5% накопленных цифровых данных, несмотря на то, что объективно существуют общеотраслевые задачи, которые можно было бы решить с помощью аналитических решений класса Big Data. Развитые IT-рынки уже имеют результаты, по которым можно оценить ожидания, связанные с накоплением и обработкой больших данных.

Одним из главных факторов, который тормозит внедрение Big Data - проектов, помимо высокой стоимости, считается **проблема выбора обрабатываемых данных**: то есть определение того, какие данные необходимо извлекать, хранить и анализировать, а какие – не принимать во внимание.

Многие представители бизнеса отмечают, что сложности при внедрении Big Data-проектов связаны с нехваткой специалистов – маркетологов и аналитиков. От качества работы сотрудников, занимающихся глубинной и предикативной аналитикой, напрямую зависит скорость возврата инвестиций в Big Data. Огромный потенциал уже существующих в организации данных часто не может быть эффективно использован самими маркетологами из-за устаревших бизнес-процессов или внутренних регламентов. Поэтому часто проекты Big Data воспринимаются бизнесом как сложные не только в реализации, но и в оценке результатов: ценности собранных данных. Специфика работы с данными требует от маркетологов и аналитиков переключения внимания с технологий и создания отчетов на решение конкретных бизнес-задач.

Для успешного принятия решения о внедрении решений Big Data компании необходимо рассчитать инвестиционный кейс и это вызывает большие трудности из-за множества неизвестных составляющих. Парадоксом аналитики в подобных случаях становится прогнозирование будущего на основе прошлого, данные о котором зачастую отсутствуют. В этом случае важным фактором является четкое планирование своих первоначальных действий:

- Во-первых, необходимо определить одну конкретную задачу бизнеса, для решения которой будут использоваться технологии Big Data, эта задача станет стержнем определения верности выбранной концепции. Необходимо сосредоточиться на сборе данных, связанных именно с этой задачей, а в ходе проверки концепции вы сможете использовать различные инструменты, процессы и методы управления, которые позволят принимать более обоснованные решения в будущем.
- Во-вторых, маловероятно, что компания без навыков и опыта аналитики данных сможет успешно реализовать проект Big Data. Необходимые знания всегда вытекают из предыдущего опыта аналитики, что является основным фактором, влияющим на качество работы с данными. Важную роль играет культура использования данных, так как часто анализ информации открывает суровую правду о бизнесе, и чтобы принять эту правду и работать с ней, необходимы выработанные методы работы с данными.
- В третьих, ценность технологий Big Data заключается в предоставлении инсайтов. Хорошие аналитики остаются дефицитом на рынке. Ими принято называть специалистов, имеющих глубокое понимание коммерческого смысла данных и знающих, как правильно их применять. Анализ данных является средством для достижения целей бизнеса, и чтобы понять ценность Big Data, необходима соответствующая модель поведения и понимание своих действий. В этом случае большие данные дадут массу полезной информации о потребителях, на основе которой можно принять полезные для бизнеса решения. [2]

Последовательность работы с Big Data состоит из сбора данных, структурирования полученной информации с помощью отчетов и дашбордов (dashboard), создания инсайтов и контекстов, а также формулирования рекомендаций к действию. Так как работа с Big Data подразумевает большие затраты на сбор данных, результат обработки которых заранее неизвестен, основной задачей является четкое понимание, для чего нужны данные, а не то, как много их есть в наличии. В этом случае сбор данных превращается в процесс получения исключительно нужной для решения конкретных задач информации.

Например, у телекоммуникационных провайдеров агрегируется огромное количество данных, в том числе о геолокации, которые постоянно пополняются. Эта информация может представлять коммерческий интерес для рекламных агентств, которые могут использовать ее для показа таргетированной и локальной рекламы, а также для ритейлеров и банков. Подобные данные могут сыграть важную роль при решении открытия торговой точки в определенной локации на основе данных о наличии мощного целевого потока людей. Есть пример измерения эффективности рекламы на outdoor-щитах в Лондоне. Сейчас охват подобной рекламы можно измерить лишь поставив возле рекламных конструкций людей со специальным устройством, подсчитывающим прохожих. По сравнению с таким видом измерения эффективности рекламы, у мобильного оператора куда больше возможностей – он точно знает местонахождение своих абонентов, ему известны их демографические характеристики, пол, возраст, семейное положение, и т.д.

На основе таких данных, в будущем открывается перспектива менять содержание рекламного сообщения, используя предпочтения конкретного человека, проходящего мимо рекламного щита. Если данные показывают, что проходящий мимо человек много путешествует, то ему можно будет показать рекламу курорта. Организаторы футбольного матча могут оценить количество болельщиков только когда те придут на матч. Но если бы они имели возможность запросить у оператора сотовой связи информацию, где посетители находились за час, день или месяц до матча, то это дало бы организаторам возможность планировать места для размещения рекламы следующих матчей. [3]

Несмотря на то, что российский рынок Big Data только начинает формироваться, отдельные проекты в этой области уже реализуются достаточно успешно. Некоторые из них успешны в области сбора данных как, например, проекты для ФНС и банка «Тинькофф Кредитные Системы», другие – в части анализа данных и практического применения его результатов: это проект Synqera.

В банке «Тинькофф Кредитные Системы» был реализован проект по внедрению платформы EMC2 Greenplum, которая является инструментом для массивно-параллельных вычислений. В течение последних лет у банка выросли требования к скорости обработки накопленной информации и анализа данных в режиме реального времени, вызванные высокими темпами роста количества пользователей кредитных карт. Банк объявил о планах расширения использования технологий Big Data, в частности для обработки неструктурированных данных и работы с корпоративной информацией, получаемой из разных источников.

В ФНС России в настоящий момент идет создание аналитического слоя федерального хранилища данных. На его основе создается единое информационное пространство и технология доступа к налоговым данным для статистической и аналитической обработки. В ходе реализации проекта выполняются работы по централизации аналитической информации с более чем 1200 источниками местного уровня ИФНС.

Еще одним интересным примером анализа больших данных в режиме реального времени является российский стартап Synqera, который разработал платформу Simplate. Решение основано на обработке больших массивов данных, программа анализирует информацию о покупателях, историю их покупок, возраст, пол и даже настроение. На кассах в сети косметических магазинов были установлены сенсорные экраны с датчиками, распознающими эмоции покупателей. Программа определяет настроение человека, анализирует информацию о нем, определяет время суток и сканирует базу скидок магазина, после чего отправляет покупателю таргетированные сообщения об акциях и специальных предложениях. Это решение повышает покупательскую лояльность и увеличивает продажи ритейлеров.

Литература.

1. Виктор Майер-Шенбергер, Кеннет Кукьер. Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим. — Изд.: Манн, Иванов и Фербер, 2013 — 260 с.
2. Френкс Б. Укрощение больших данных. Как извлекать знания из массивов информации с помощью глубокой аналитики. — Изд.: Технология развития ООО, 2014 — 352 с.
3. Работа с Big Data: основные области и возможности. — [Электронный ресурс]. — Режим доступа: http://www.marketing.spb.ru/lib-research/methods/Big_Data.htm